



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE CÓMPUTO

NOMBRES:

PADILLA MATIAS CRISTIAN MICHEL

SAUCILLO GONZÁLEZ JESSE OBED

GRUPO:

4BM1

TRABAJO:

Practica 8

"Algoritmo K-NN"

MATERIA:

Fundamentos de Inteligencia Artificial

FECHA

4 de diciembre del 2023

INSTITUTO POLITÉCNICO NACIONAL



ESCOM

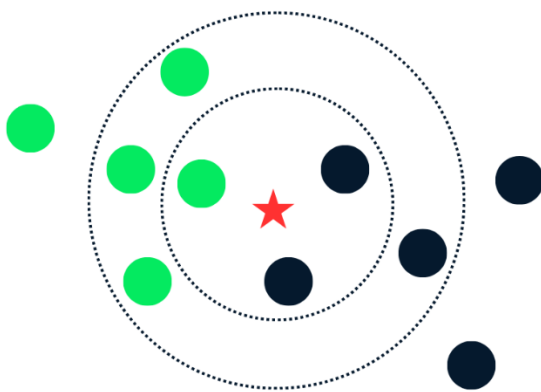
PRACTICA 5. Algoritmo K-NN

RESUMEN

La práctica se centra en la implementación del algoritmo K-NN (K-Nearest Neighbors) para clasificar países en categorías de ingresos según la esperanza de vida al nacer, el PIB per cápita y las emisiones de CO2 per cápita. Se utiliza un conjunto de entrenamiento y prueba seleccionado estratégicamente para representar equitativamente las cuatro clases de ingresos: 'High', 'Upper Middle', 'Lower Middle' y 'Low'. La práctica busca explorar el rendimiento del algoritmo en la clasificación de nuevos datos.

INTRODUCCIÓN

El algoritmo K-NN pertenece a la categoría de los algoritmos de "aprendizaje perezoso" (lazy learning), lo que significa que la generalización de los datos de entrenamiento se retrasa hasta que se hace una consulta para realizar una predicción. A pesar de su nombre, los clasificadores perezosos son muy útiles cuando se trabaja con grandes conjuntos de datos que se actualizan con frecuencia o en tiempo real, y no se puede estar entrenando un algoritmo cada vez que se agregan instancias o cambian los valores de los atributos



El algoritmo K-NN funciona encontrando los K vecinos más cercanos a un punto de datos dado en función de una métrica de distancia, como la distancia euclidiana. Para tareas de clasificación, el algoritmo asigna la etiqueta que es más frecuente entre los K vecinos más cercanos. Para tareas de regresión, el algoritmo calcula el promedio o el

promedio ponderado de los valores de los K vecinos más cercanos y lo asigna como el valor predicho para el punto de datos

DESARROLLO

Se desarrollo la practica en el lenguaje Python en donde implementamos el algoritmo K-NN utilizando la biblioteca scikit-learn en Python. Despues, se carga el conjunto de datos desde un archivo CSV que contiene información sobre la esperanza de vida, el PIB per cápita, las emisiones de CO2 y la clase de ingresos de diversos países.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

df = pd.read_csv('países2.csv')

X = df[['Esperanza de vida al nacer', 'PIB per cápita $US', 'Emisiones de CO2 per cápita']]
y = df['Clase']
```

Se selecciona un subconjunto de 100 países para el entrenamiento y 20 para las pruebas, asegurando una representación equitativa de las clases de ingresos. Ahí se normalizan los datos y se dividen en conjuntos de entrenamiento y prueba.

```
Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=20, random_state=42)
```

Procedemos con el entrenamiento del modelo K-NN con el conjunto de entrenamiento y se realizan predicciones sobre el conjunto de prueba para diferentes valores de K.

```
Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=20, random_state=42)

knn = KNeighborsClassifier(n_neighbors=10)
knn.fit(Xtrain, ytrain)

prediccion = knn.predict(Xtest)
```

Por último, se visualiza el resultado mediante una gráfica de dispersión tridimensional que muestra la distribución de los países en función de la esperanza de vida, el PIB per cápita y las emisiones de CO2.

```

ax.scatter(Xtrain[ytrain == 'Low income']['Esperanza de vida al nacer'],
          Xtrain[ytrain == 'Low income']['PIB per cápita $US'],
          Xtrain[ytrain == 'Low income']['Emisiones de CO2 per cápita'],
          c="purple", Label='Low income')

ax.scatter(Xtrain[ytrain == 'Lower middle income']['Esperanza de vida al nacer'],
          Xtrain[ytrain == 'Lower middle income']['PIB per cápita $US'],
          Xtrain[ytrain == 'Lower middle income']['Emisiones de CO2 per cápita'],
          c="violet", Label='Lower middle income')

ax.scatter(Xtrain[ytrain == 'Upper middle income']['Esperanza de vida al nacer'],
          Xtrain[ytrain == 'Upper middle income']['PIB per cápita $US'],
          Xtrain[ytrain == 'Upper middle income']['Emisiones de CO2 per cápita'],
          c="palegreen", Label='Upper middle income')

ax.scatter(Xtrain[ytrain == 'High income']['Esperanza de vida al nacer'],
          Xtrain[ytrain == 'High income']['PIB per cápita $US'],
          Xtrain[ytrain == 'High income']['Emisiones de CO2 per cápita'],
          c="greenyellow", Label='High income')

ax.scatter(Xtest['Esperanza de vida al nacer'],
          Xtest['PIB per cápita $US'],
          Xtest['Emisiones de CO2 per cápita'],
          c="black", marker='x', Label = 'Países introducidos')
ax.legend()
plt.show()

```

Los datos extraídos para el entrenamiento y el testeo correspondientes al archivo 'países.csv' son los siguientes:

Nombre	Esperanza de vida al nacer	PIB per capita \$US	Emisiones de CO2 per capita	Clase
Afghanistan	61.982	368.754614	6075.23	Low income
Albania	76.463	6492.87201	4829.2	Upper middle income
Algeria	76.377	3690.62788	171249.6	Lower middle income
Angola	61.643	1953.53376	25212.9	Lower middle income
Antigua and Barbuda	78.497	15781.3957	516.1	High income
Argentina	75.39	10636.1202	168095.3	Upper middle income
Armenia	72.043	4966.51347	6169.6	Upper middle income
Australia	83.3	60443.1092	386526.6	High income
Austria	81.2390244	53637.7057	64766.3	High income
Azerbaijan	69.366	5387.99798	35508	Upper middle income
Bahamas, The	71.598	27478.3896	2836.1	High income
Bahrain	78.76	26562.9691	33256.7	High income
Barbados	77.571	17225.4623	1249.97	High income
Belarus	72.3706829	7302.25779	57672.2	Upper middle income

Belgium	81.8902439	51247.0144	93008.5	High income
Belize	70.47	6228.26731	639.8	Upper middle income
Benin	59.821	1319.155	7300.3	Lower middle income
Bhutan	71.815	3266.36491	1053.8	Lower middle income
Bolivia	63.63	3345.19659	22340.7	Lower middle income
Bosnia and Herzegovina	75.3	7143.31055	21066.9	Upper middle income
Botswana	61.141	6805.22127	7252.5	Upper middle income
Brazil	72.75	7507.16097	434297.9	Upper middle income
Brunei Darussalam	74.642	31449.0766	6986.2	High income
Bulgaria	71.5146342	12221.4966	39143.1	Upper middle income
Burkina Faso	59.27	893.077156	5003.8	Low income
Burundi	61.663	221.477676	720.72	Low income
Cabo Verde	74.052	3293.23305	645.3	Lower middle income
Cambodia	69.584	1625.23502	16177.4	Lower middle income
Cameroon	60.333	1666.93273	9429.1	Lower middle income
Canada	82.5966098	51987.9391	580212.8	High income
Central African Republic	53.895	461.13749	242.6	Low income
Chad	52.525	685.690284	2247.7	Low income
Chile	78.944	16265.096	91369.4	High income
China	78.211	12556.3331	10707220	Upper middle income
Colombia	72.83	6104.13671	81039.1	Upper middle income
Comoros	63.417	1577.47103	320.2	Lower middle income
Congo, Dem. Rep.	59.193	577.209215	3209.8	Low income
Congo, Rep.	63.519	2290.38289	7256.9	Lower middle income
Costa Rica	77.023	12472.4437	8238.1	Upper middle income
Cote d'Ivoire	58.598	2549.0413	10833.4	Lower middle income
Croatia	76.4243902	17685.3253	16523.3	High income
Cyprus	81.203	31551.8164	7193.2	High income

Czechia	77.3731707	26821.2452	96285.2	High income
Denmark	81.4048781	68007.7567	29696.2	High income
Djibouti	62.305	3150.43673	420.02	Lower middle income
Dominica	72.814	7653.17187	167.8	Upper middle income
Dominican Republic	72.615	8476.75216	27233.3	Upper middle income
Egypt, Arab Rep.	70.221	3698.83498	249367.1	Lower middle income
El Salvador	70.748	4551.18466	7797.8	Upper middle income
Equatorial Guinea	60.594	7506.66752	5351.8	Upper middle income
Estonia	76.7414634	27943.7012	10175.4	High income
Eswatini	57.066	3978.40353	963	Lower middle income
Ethiopia	64.975	925.077428	18359.3	Low income
Fiji	67.114	4646.61272	1632.37	Upper middle income
Finland	81.9341463	53654.7503	40706.8	High income
France	82.3243902	43658.979	300515.6	High income
Gabon	65.821	8635.32569	5245.5	Upper middle income
Gambia, The	62.083	772.152395	577.4	Low income
Georgia	71.694	5023.27438	10113.2	Upper middle income
Germany	80.9009756	51203.5545	657401	High income
Ghana	63.795	2363.2993	20040.6	Lower middle income
Greece	80.1829268	20192.5963	59996.8	High income
Grenada	74.936	9010.57225	325	Upper middle income
Guatemala	69.237	5025.54229	19313	Upper middle income
Guinea	58.892	1189.176	3954.6	Lower middle income
Guinea-Bissau	59.652	795.118569	330.8	Low income
Guyana	65.673	9998.54431	2846.2	High income
Haiti	63.192	1829.59304	3317.34	Lower middle income
Honduras	70.123	2771.71746	10219.6	Lower middle income
Hungary	74.4658537	18728.1219	46382.3	High income

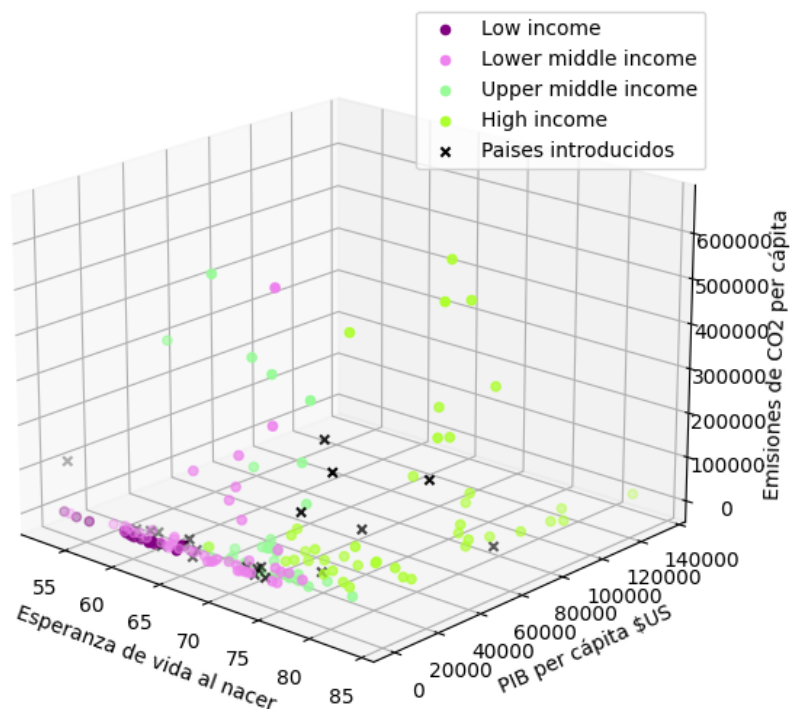
India	67.24	2256.59041	2456301	Lower middle income
Indonesia	67.57	4332.70928	619840.7	Upper middle income
Iran, Islamic Rep.	73.875	4091.20919	630014.4	Lower middle income
Ireland	82.102439	100172.079	35750.9	High income
Israel	82.5	52170.7119	62650.3	High income
Italy	82.795122	35657.4976	317238.6	High income
Jamaica	70.5	5183.58131	8394.4	Upper middle income
Japan	84.4456098	39312.6604	1081572	High income
Jordan	74.256	4103.25897	24627.5	Lower middle income
Kazakhstan	70.23	10373.7898	212107.9	Upper middle income
Kenya	61.427	2081.79986	22281.7	Lower middle income
Kiribati	67.417	1606.46251	86.9	Lower middle income
Korea, Rep.	83.5268293	34997.7816	610788.5	High income
Kyrgyz Republic	71.9	1276.70037	10046.7	Lower middle income
Latvia	73.2829268	21148.1629	7569.2	High income
Lebanon	75.047	4136.14635	27946.5	Lower middle income
Lesotho	53.062	1094.09819	769.8	Lower middle income
Liberia	60.747	675.663186	1182	Low income
Libya	71.911	6357.19555	56795.7	Upper middle income
Lithuania	74.3390244	23723.3403	11730.3	High income
Luxembourg	82.7487805	133590.147	9489.2	High income
Madagascar	64.485	500.511032	4120.47	Low income
Malawi	62.904	634.83566	1453.7	Low income
Malaysia	74.884	11109.2618	253266.5	Upper middle income
Maldives	79.918	10366.2934	2109.6	Upper middle income
Mali	58.941	873.794862	5833.8	Low income
Marshall Islands	65.269	6172.14507	179.5	Upper middle income
Mauritania	64.364	2166.0468	3949.1	Lower middle income

Mauritius	73.6802439	9106.2372	4173.7	Upper middle income
Mexico	70.213	10045.6805	449266.1	Upper middle income
Micronesia, Fed. Sts.	70.71	3571.33677	175.2	Lower middle income
Moldova	68.846	5230.66173	8859.4	Upper middle income
Mongolia	70.975	4566.14015	23074.1	Lower middle income
Montenegro	73.8243902	9465.704	2601.1	Upper middle income
Morocco	74.042	3795.38013	71479	Lower middle income
Mozambique	59.325	491.839113	7495.1	Low income
Myanmar	65.672	1209.92694	36718.2	Lower middle income
Namibia	59.269	4865.55777	4216	Upper middle income
Nepal	68.45	1208.21853	13448.2	Lower middle income
Netherlands	81.4609756	57767.8788	146341	High income
Nicaragua	73.837	2045.5354	5236.1	Lower middle income
Niger	61.576	590.629455	2148.95	Low income
Nigeria	52.676	2065.74907	115278.1	Lower middle income
North Macedonia	74.5368293	6694.64113	8302.6	Upper middle income
Norway	83.1634146	89154.2761	35946.4	High income
Oman	72.541	19509.4665	76026.8	High income
Pakistan	66.098	1505.01019	190570.8	Lower middle income
Panama	76.223	14617.6045	13344.2	High income
Papua New Guinea	65.351	2672.94579	7582	Lower middle income
Paraguay	70.262	5891.49997	8213.7	Upper middle income
Peru	72.377	6621.57434	56746.3	Upper middle income
Philippines	69.266	3460.53096	145418.6	Lower middle income
Poland	75.602439	17999.91	295132.2	High income
Portugal	81.0731707	24567.5093	44638.6	High income
Qatar	79.272	66838.3574	91967.4	High income
Romania	72.9609756	14858.2294	73948.3	High income

Russian Federation	69.3612195	12194.7773	1703590	Upper middle income
Rwanda	66.072	822.347989	1332.816	Low income
Samoa	72.767	3857.31846	295.4	Lower middle income
Sao Tome and Principe	67.591	2360.54355	145.4	Lower middle income
Saudi Arabia	76.936	23185.8708	523778.5	High income
Senegal	67.093	1636.89321	10616.6	Lower middle income
Serbia	72.7307317	9230.17832	45946.1	Upper middle income
Sierra Leone	60.062	480.039211	898.9	Low income
Singapore	83.4414634	72794.003	47378.6	High income
Slovenia	80.8756098	29291.4006	13599.8	High income
Solomon Islands	70.348	2304.84457	357.5	Lower middle income
Somalia	55.28	446.98156	693.7	Low income
South Africa	62.341	7055.04478	439640	Upper middle income
Spain	83.1780488	30103.5137	239977.8	High income
Sri Lanka	76.399	4013.68766	23782.4	Lower middle income
St. Kitts and Nevis	71.682	18082.6102	248.8	High income
St. Lucia	71.111	9414.22623	535.7	Upper middle income
St. Vincent and the Grenadines	69.629	8666.38704	256.1	Upper middle income
Sudan	65.267	751.82135	20622.8	Low income
Suriname	70.274	4869.13423	2648.93	Upper middle income
Sweden	83.1560976	61028.7381	35001.7	High income
Switzerland	83.8512195	91991.6005	37376.3	High income
Tajikistan	71.594	897.04751	9417.6	Lower middle income
Tanzania	66.201	1099.2876	12454	Lower middle income
Thailand	78.715	7066.19055	267089.1	Upper middle income
Timor-Leste	67.737	2741.39393	615.4	Lower middle income
Togo	61.619	973.206129	2365.4	Low income
Tonga	70.986	4426.00064	162.9	Upper middle income
Trinidad and Tobago	72.971	16032.5028	17186.7	High income

Tunisia	73.772	3807.13915	29908.4	Lower middle income
Turkiye	76.032	9661.23598	396842.9	Upper middle income
Tuvalu	64.547	5632.00299	11	Upper middle income
Uganda	62.705	883.892032	5862.1	Low income
Ukraine	69.6478049	4835.57178	174731.8	Lower middle income
United Arab Emirates	78.71	44315.5542	188858.6	High income
United Kingdom	80.7	46510.2828	348921.9	High income
United States	76.3292683	70248.629	4817719	High income
Uruguay	75.436	17313.1884	6490.1	High income
Uzbekistan	70.862	1983.06472	116707.2	Lower middle income
Vanuatu	70.449	2996.62106	210.8	Lower middle income
Vietnam	73.618	3756.48912	336491.8	Lower middle income
Zambia	61.223	1137.34363	6803.9	Lower middle income
Zimbabwe	59.253	1773.92041	11759.9	Lower middle income

A continuación, la gráfica de dispersión con los datos de entrenamiento y los datos de testeo:



Pruebas con distintos valores en k

Para un K = 5, los resultados son:

```
La categoría predicha es:  
['Lower middle income' 'Upper middle income' 'Low income' 'High income'  
 'Upper middle income' 'High income' 'Upper middle income'  
 'Lower middle income' 'Upper middle income' 'Upper middle income'  
 'Lower middle income' 'Upper middle income' 'High income'  
 'Lower middle income' 'Lower middle income' 'High income' 'Low income'  
 'High income' 'Lower middle income' 'High income']  
Cantidad de High income en los paises de prueba: 6  
Cantidad de Upper middle income en los paises de prueba: 6  
Cantidad de Lower middle income en los paises de prueba: 6  
Cantidad de Low income en los paises de prueba: 2
```

Para un K = 10, los resultados son:

```
La categoría predicha es:  
['Lower middle income' 'Lower middle income' 'Low income' 'High income'  
 'Lower middle income' 'High income' 'Upper middle income'  
 'Lower middle income' 'Upper middle income' 'Upper middle income'  
 'High income' 'Lower middle income' 'High income' 'Lower middle income'  
 'Lower middle income' 'High income' 'Low income' 'High income'  
 'Lower middle income' 'High income']  
Cantidad de High income en los paises de prueba: 7  
Cantidad de Upper middle income en los paises de prueba: 3  
Cantidad de Lower middle income en los paises de prueba: 8  
Cantidad de Low income en los paises de prueba: 2
```

Para un K = 15, los resultados son:

```
La categoría predicha es:  
['Lower middle income' 'Lower middle income' 'Low income' 'High income'  
 'Lower middle income' 'High income' 'Upper middle income'  
 'Lower middle income' 'Upper middle income' 'Upper middle income'  
 'High income' 'Lower middle income' 'Upper middle income'  
 'Lower middle income' 'High income' 'High income' 'Low income'  
 'High income' 'Lower middle income' 'High income']  
Cantidad de High income en los paises de prueba: 7  
Cantidad de Upper middle income en los paises de prueba: 4  
Cantidad de Lower middle income en los paises de prueba: 7  
Cantidad de Low income en los paises de prueba: 2
```

Para un $K = 20$, los resultados son:

```
La categoría predicha es:  
['Lower middle income' 'Lower middle income' 'Low income' 'High income'  
 'Lower middle income' 'High income' 'Lower middle income'  
 'Lower middle income' 'Upper middle income' 'Upper middle income'  
 'High income' 'Lower middle income' 'High income' 'Lower middle income'  
 'High income' 'High income' 'Low income' 'High income'  
 'Lower middle income' 'High income']  
Cantidad de High income en los países de prueba: 8  
Cantidad de Upper middle income en los países de prueba: 2  
Cantidad de Lower middle income en los países de prueba: 8  
Cantidad de Low income en los países de prueba: 2
```

CONCLUSIONES

El algoritmo K-NN es útil en la clasificación de conjuntos de datos como de países según sus características socioeconómicas. La elección de K puede influir en el rendimiento del modelo, y se recomienda seleccionar el valor adecuado de K mediante la ejecución del algoritmo varias veces con diferentes valores de K y eligiendo el que reduzca el número de errores que se encuentran, manteniendo la capacidad del algoritmo para hacer predicciones precisas cuando se le da datos que no ha visto antes.

El algoritmo K-NN es simple y fácil de implementar, y se adapta fácilmente a la adición de nuevas muestras de entrenamiento, ya que todos los datos de entrenamiento se almacenan en memoria. Aunque el algoritmo tiene fortalezas y debilidades, dependiendo del proyecto y la aplicación, puede ser una buena opción para clasificar de datos que requieren de una actualización frecuente o en tiempo real.

BIBLIOGRAFÍA

Harrison, O. (2019, 14 julio). Machine Learning basics with the K-Nearest Neighbors

Algorithm. *Medium*. <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

GeeksforGeeks. (2023, 9 noviembre). *K Nearest neighbor KNN algorithm*.

<https://www.geeksforgeeks.org/k-nearest-neighbours/>